A Connection between Paired Data Analysis and Regression Analysis for Estimating Sales Adjustments

*Joseph B. Lipscomb**

*J. Brian Gray***

Abstract. **The two methods most often recommended for obtaining market-derived adjustments utilized in the sales comparison approach to appraisal are Paired Data Analysis and Multiple Regression Analysis. These approaches are viewed as competing alternatives, with advocates and detractors for each. The main purpose of this paper is to demonstrate that these two alternatives to estimating sales adjustments are equivalent under certain circumstances. This point of equivalence may prove to be a useful starting place for improving our understanding of the differences between and similarities of the two methods. After explaining the data requirements of each method, we provide a set of sufficient conditions under which the two methods produce identical adjustment estimates. We finish with a discussion of the relative advantages and disadvantages of these two methods in estimating sale comparison adjustments.**

Introduction

In the appraisal literature, Paired Data Analysis (PDA) and Multiple Regression Analysis (MRA) are the two methods most often recommended for obtaining market-derived grid adjustments used in the sales comparison approach. Most researchers and some mass appraisal institutions rely on regression coefficients from MRA for adjustments (Cannaday, 1989; Colwell, Cannaday and Wu, 1983; Kang and Reichert, 1991; Lipscomb and Gray, 1990; Tchira, 1979; Vandell, 1992). Fee appraisers typically rely on PDA and are skeptical of MRA as a substitute method for estimating the value of adjustments. In this article, we provide a set of sufficient data conditions under which the two methods are equivalent and thus produce identical adjustment estimates.

In practice, PDA and MRA differ in two fundamental respects. The first difference has to do with the number of sale observations typically used and whether or not the sales are paired. In principle, PDA requires pairs of property sales matched in every respect except the *primary attribute* for which an adjustment (i.e., the *primary adjustment*) is desired. Practitioners admit that perfect pairs (or even well-matched pairs) are hard to find. Although there is no established standard for the minimum number of paired sales required to estimate an adjustment value, examples of adjustments based on as few as three pairs of sales are common.[1] MRA does not require matching per se, making data selection simpler and usable observations more readily available.[2] The minimum number

*M. J. Neeley School of Business, Texas Christian University, Fort Worth, Texas 76129.
**Department of Management Science and Statistics, University of Alabama, Tuscaloosa, Alabama 35487.
Date Revised—January 1995; Accepted—January 1995.

of observations required for MRA is two more than the number of explanatory variables (to allow for an estimate of the error variance). All additional observations that add information without introducing unacceptable bias should be, and typically are, included in the MRA data set.

The second difference between the two methods is the way in which they account for differences in attributes other than the primary attribute that affect value (i.e., *secondary attributes*). PDA accomplishes this by matching pairs of properties as closely as possible on all determinants of selling price except the primary attribute. However, in PDA, *secondary adjustments* are usually needed to correct for leftover differences in a matched pair. If paired sales differ in some attribute other than the primary attribute, a secondary adjustment corrects the sale price of one sale by adding (or subtracting) the net value of this unwanted difference.[3] With PDA, secondary adjustment values can be supplied by the analyst from one or more sources, including other paired data, cost data, survey data, and regression estimates. As a practical matter, with PDA there are few constraints on information sources or methods of estimating secondary adjustments.

In contrast, the MRA model uses regression coefficients to account for the attributes that affect value. In effect, MRA decomposes sales prices into components due to or explained by these attributes, including the primary attribute. While it is also possible to bring outside sources of information into the MRA procedure (see, e.g., Gilley and Pace, 1990), it is not as easily accomplished as with PDA. In this article, we assume that only the information in the MRA data set is used in computing corrections for secondary attribute differences.[4]

To summarize, PDA requires a paired data set while MRA makes no similar requirement. Because these two methods have different data structure requirements and deal with the contributory value of secondary attributes in dissimilar fashions, any relationship between the two methods is not readily apparent. In this paper, we show that if the information available to both methods is limited to a paired data set and secondary adjustments for the PDA method are regression coefficients estimated internally from the paired data set, then the MRA and PDA methods are equivalent.

The remainder of this paper is organized as follows: Section two is a description of adjustment estimation through the methods of Paired Data Analysis (PDA) and Multiple Regression Analysis (MRA). Section three establishes sufficient conditions for the equivalence of the two methods. The fourth section contains a discussion of the relative merits of MRA and PDA for estimating adjustments, including the role of sources of information external to the techniques themselves. Finally, section five provides conclusions.

Adjustments from Paired Data Analysis and Multiple Regression Analysis

Suppose that a market value adjustment is sought for a particular attribute. In any form of paired data analysis, a pair of sales is identified such that the properties in the pair are similar in all major respects (specifically in location, market condition, financing, and all important physical attributes) except for this primary attribute. One of the properties in the pair has the primary attribute and the other does not. The sale price difference between the properties in this otherwise perfectly matched pair is attributed to the presence of the primary attribute in one of the properties.

Each observation in the data set will be referred to as a sale. We assume that the sale price can be described by the linear model

$$Y = \beta_0 + \sum_{k=1}^{p} \beta_k X_k + \beta_W W + \varepsilon, \tag{1}$$

with sale price Y as the response variable, secondary attributes X_1, X_2, . . ., X_p as explanatory variables, and the indicator (or dummy) variable W representing the primary attribute, where W is 1 if the sale has the primary attribute, and 0 otherwise. (Note that we are assuming that the effect of the primary attribute on sale price is additive and constant with respect to other attributes. Colwell, Cannaday and Wu (1983) describe alternative formulations.) The random errors ε are typically assumed to be independent and identically distributed with constant, but unknown, variance.

Given a sales data set of interest, the process of constructing a paired data set begins by identifying n sales with the primary attribute, where $n \geq {}^{1}/_{2}(p+3)$ to allow for an estimate of the error variance. From the remaining sales in the database without the primary attribute, the best match for each of these n sales is found (without replacement).[5] The total number of sales in the paired data set is then $2n$, half with and half without the primary attribute. The analysis of the paired data set from this point forward differs for the two adjustment methods, but as we shall see later, the results are the same under certain conditions.

MRA proceeds by fitting model (1) to the $2n$ sales in the paired data set using least squares regression. (Recall that in this section we are restricting MRA to the paired data set.) The fitted model is given by

$$\hat{Y} = \hat{\beta}_0 + \sum_{k=1}^{p} \hat{\beta}_k X_k + \hat{\beta}_W W, \tag{2}$$

where $\hat{\beta}_1$, $\hat{\beta}_2$, . . ., and $\hat{\beta}_p$ are the least squares estimates of the model parameters representing the marginal dollar values of the secondary attributes X_1, X_2, . . ., X_p, and $\hat{\beta}_W$ is the primary adjustment estimate (for the primary attribute).

In theory, PDA assumes that the pairs of sales in the paired data set are perfectly matched in all respects except the primary attribute. In principle, the dollar adjustment amount for the primary attribute is estimated by the average of the price differences in the n pairs of sales. In practice, however, the perfect match requirement is rarely met.[6] Consequently, most pairs require adjustments to the selling price of one of the sales in each pair to control for unwanted differences in secondary attributes. After these secondary adjustments have been made, a pair of sales is considered to be perfectly matched except on the primary attribute. Any remaining difference in adjusted selling prices is taken to be an estimate of the adjustment for the primary attribute. The price differences from the n pairs are averaged to produce the final primary adjustment estimate.

Secondary adjustment values for the unwanted secondary attribute differences can be estimated by the regression coefficients $\hat{\beta}_1$, $\hat{\beta}_2$, . . ., and $\hat{\beta}_p$ in equation (2). As we will see in the next section, these estimates of the secondary adjustment amounts have useful properties. For now, let Y_{j1} represent the price of the property with the primary attribute in the jth pair ($W=1$) and Y_{j0} represent the price of the property without the primary attribute in the jth pair ($W=0$). Then the price difference between the two sales in the jth pair is given by $Y'_j = Y_{j1} - Y_{j0}$. Similarly, let $X'_{ij} = Y_{ij1} - Y_{ij0}$ represent the difference in the ith explanatory variable values for the two sales in the jth pair. The secondary adjustment to the price difference in the jth pair of sales due to the lack of matching in the ith secondary attribute is given by $\hat{\beta}_i X'_{ij}$, and the total net value of all secondary adjustments to the price difference in the jth pair is computed as

$$Adj_j = \sum_{k=1}^{p} \hat{\beta}_k X'_{kj}. \tag{3}$$

The adjustment amount for the primary attribute indicated by the jth pair is then given by

$$AdjY'_j = Y'_j - Adj_j. \tag{4}$$

A more accurate estimate of the primary adjustment amount is found by averaging the adjusted differences of all n pairs to obtain

$$\overline{AdjY'} = \frac{1}{n} \sum_{j=1}^{n} AdjY'_j. \tag{5}$$

This is the primary adjustment estimate resulting from the PDA method.

In the next section, we establish the equivalence between the PDA and the MRA methods by showing that $\overline{AdjY'} = \hat{\beta}_W$ when both methods are restricted to the paired data set and secondary adjustments for the PDA method are determined from MRA regression coefficients.

An Equivalence between Paired Data Analysis and Multiple Regression Analysis

Equation (4), which gives the primary adjustment amount indicated by the jth pair, can be rewritten in an expanded form by substituting the definition of the secondary adjustment from equation (3):

$$AdjY'_j = Y'_j - \sum_{k=1}^{p} \hat{\beta}_k X'_{kj}. \tag{6}$$

By substituting $Y'_j = Y_{j1} - Y_{j0}$ and $X'_{ij} = Y_{ij1} - Y_{ij0}$, multiplying the terms and rearranging, the result is

$$AdjY'_j = Y_{j1} - Y_{j0} - \sum_{k=1}^{p} \hat{\beta}_k (X_{kj1} - X_{kj0}) \tag{7}$$

$$= \left[Y_{j1} - \sum_{k=1}^{p} \hat{\beta}_k X_{kj1} \right] - \left[Y_{j0} - \sum_{k=1}^{p} \hat{\beta}_k X_{kj0} \right]. \tag{8}$$

Using the MRA fitted equation (2), the terms in equation (8) can be simplified to obtain

$$AdjY_j' = \left[Y_{j1} - \left(\hat{Y}_{j1} - \hat{\beta}_0 - \hat{\beta}_W \cdot 1\right)\right] - \left[Y_{j0} - \left(\hat{Y}_{j0} - \hat{\beta}_0 - \hat{\beta}_W \cdot 0\right)\right] \tag{9}$$

$$= \hat{\beta}_W + \left[Y_{j1} - \hat{Y}_{j1}\right] - \left[Y_{j0} - \hat{Y}_{j0}\right] \tag{10}$$

$$= \hat{\beta}_W + e_{j1} - e_{j0}, \tag{11}$$

where e_{j1} and e_{j0} are the residuals for the two sales in the jth pair from the regression in equation (2). If the $AdjY_j'$ values are averaged over all n pairs, the desired result

$$\overline{AdjY'} = \frac{1}{n}\sum_{j=1}^{n} AdjY_j' = \frac{1}{n}\sum_{j=1}^{n}(\hat{\beta}_W + e_{j1} - e_{j0}) = \hat{\beta}_W \tag{12}$$

follows because

$$\sum_{j=1}^{n} e_{j1} = \sum_{j=1}^{n} e_{j0} = 0. \tag{13}$$

The fact that equation (13) is true can be established from well-known results from least squares regression analysis. First, the sum of all residuals from a least squares regression is zero (see, e.g., Hamilton, 1992, pp. 112–13). In our case, this implies:

$$\sum_{j=1}^{n} e_{j1} + \sum_{j=1}^{n} e_{j0} = 0. \tag{14}$$

Secondly, the least squares residuals are uncorrelated with the explanatory variables (see, e.g., Hamilton, 1992, pp. 112–13). This implies that the correlation (and hence, the covariance) of the residuals with the dummy variable W is zero. Since the mean of the residuals is zero, it follows that the sum of cross-products between the residuals and the values of W is zero:

$$0 = \sum_{j=1}^{n} W_j e_{j1} + \sum_{j=1}^{n} W_j e_{j0} = \sum_{j=1}^{n} W_j e_{j1}. \tag{15}$$

The last equality follows from the fact that $W=0$ for sales without the primary attribute. Taken together with equation (14), this implies that the residuals associated with $W=1$ sum to zero. Therefore, equation (13), and hence equation (12), follows. As a result, PDA and MRA are equivalent under the following sufficient *conditions for equivalence*:

1. The set of attribute variables used to describe the sales, and also used to match on for establishing the pairs using PDA, are the same set of attributes that make up the explanatory variables in the regression analysis.

2. The effects of the attributes (explanatory variables) used in PDA (MRA) are linearly additive in their contribution to the total value of the sales observations.
3. Both methods use the same data set, in particular, the MRA data set must consist of the same sale observations and attributes as used for PDA.
4. The secondary adjustment values used in PDA are provided from the regression coefficients from MRA based on the PDA data set.

Lipscomb and Gray (1990) provide a detailed numerical example demonstrating this equivalence.

Discussion

PDA and MRA are treated in the appraisal literature as competing, alternative methods for estimating adjustment values. We have shown that these seemingly different methods are equivalent when the conditions for equivalence outlined above are met. Although the data conditions for equivalence would not likely produce the best available data set for either PDA or MRA in an applications context, they provide a useful theoretical starting point for understanding the differences between the two methods. We point out that the data conditions for equivalence represent the intersection between the data sets normally used by PDA and MRA (i.e., only paired data are used and, at the same time, all secondary adjustments required for the PDA method are supplied from regression coefficients estimated internally from the paired data set itself). By recognizing how data conditions normally employed in practice differ from these data conditions for equivalence, we may gain useful insight into the nature and degree of the differences that exist between the two adjustment methods in practice. In that regard, we offer the following comparison based on two significant differences in the data sets normally used with each method: (1) the many observations of MRA versus the paired data of PDA, and (2) the methods and data used to deal with imperfect matches in PDA versus MRA's handling of attribute differences.

Many Observations of MRA versus the Paired Data of PDA

In practice, MRA can take advantage of the information provided by all available observations while PDA is limited to those observations that provide "good" matches. Using more observations tends to reduce the standard errors of the estimated coefficients. On the other hand, there are practical limits to the notion that more observations in a regression result in better estimates (Gau and Kohlhepp, 1978). In most applications, statistical models (e.g., the linear regression model) are only considered to be reasonable approximations to reality over a relatively small region of the data space. Any observations that are relatively remote in the data set may introduce bias into the estimates.

While there are clearly advantages to using many observations, limiting the data to paired sales has its own benefits. PDA does not assume an explicit form of the model as does MRA (Colwell, Cannaday and Wu, 1983). By matching properties as closely as possible, the effects of the matched variables are considered without having to specify the form of their impact on the sale price. In the event of perfect matches, PDA is essentially a model-free approach (assuming that the true effect of the primary attribute on sale price is additive). Even in those cases where the matches are not exact, the partial matching in a well-matched sample accounts for most of the impact of these imperfectly matched secondary attributes on the sale price. The bias introduced by adjusting out the remaining small differences may be smaller in many instances than the bias due to incorrect model specification that might be introduced by MRA.

Dealing with Imperfect Matches and the Sources of Secondary Adjustments

In practice, users of PDA normally go outside of the paired data set for secondary adjustment information. Restricting secondary adjustments for PDA to only the data contained in the paired data set is probably too limiting in view of the small data sets that are normally available in practice. The alternative sources are cost and survey data, adjustments derived from a different set of paired data, and regression coefficients from a different data set. While both MRA and PDA have the ability to incorporate non-market sources of information (Gilley and Pace, 1990), if market-derived adjustments are required, sources of information such as cost or survey data are inapplicable. Secondary adjustments derived from a different paired data set may not be appropriate for the data at hand. Additionally, if this paired data adjustment is being calculated to serve as a secondary adjustment for the original set of paired data, then where will we obtain the secondary adjustments for this paired data set? Presumably it will be from a third set of paired data, which logically will require a fourth set and so on. This points out a problem in using paired data as secondary adjustments for other matched pairs. The final alternative is regression coefficients from an external data set. This raises the following question: If regression coefficients provide the best source of secondary adjustments, why would they not also be the best source of the primary adjustment?

With MRA, secondary adjustments are not needed. Secondary attribute differences are compensated for simultaneously through the normal course of calculating regression coefficients. All the information used is drawn from the regression data set, and MRA estimates have the virtue of being market derived.

Conclusion

In conclusion, MRA does well when many observations are available. The ability of MRA to use large data sets reduces the standard error of the coefficient estimates. This is offset by the potential introduction of bias due to model specification error and outlying data. And finally, the MRA process provides a purely market-derived solution for adjustment values.

PDA can be employed when a sufficient number of paired sales is available. It has the advantage of being virtually model-free, but it is burdened by the need for secondary adjustments from some outside source which frequently includes non-market information.

We have shown when and why PDA and MRA are equivalent. We also discussed how they differ in practice. However, we do not know how much their results differ in practice and which method has the advantage under different real world circumstances. These distinctions can best be learned by comparing the two methods using real data and a known primary attribute value.

Notes

[1]In a non-regression setting, Vandell (1992) has shown that, in choosing comparables for the sales comparison approach, it is always desirable to consider more comparables as long as their adjusted value estimates are optimally weighted in the final value estimate.

[2]Even with MRA, some selection of sale observations is customary although it is not strictly necessary, in principle. Restricting sale observations to properties with similar attributes and market characteristics, implicitly a form of matching, helps in reducing bias in the estimates caused by model inadequacy (e.g., nonlinearity over broader ranges of characteristics). However, the larger the number of observations included in the estimate, the smaller the variance of the estimates. Consequently, there is a trade-off between minimizing the variance of the estimate and reducing the bias due to model inadequacy. The "optimum" number of observations depends on the adequacy of the model, the variation in Y, and the design of the explanatory variable values.

[3]An example will help clarify this point and establish the terminology. If an adjustment value is needed for the existence (absence) of a fireplace, pairs of sales that include one house with a fireplace and a second house without a fireplace are required. In this case, fireplace is the *primary attribute.* Ideally, each house in a pair would be identical to its mate in every attribute except fireplace. Applying PDA, the difference between the sale prices of the two houses in each pair provides an estimate of the *primary adjustment* value for a fireplace. If the only difference is in the primary attribute, the pair is called a *pure pair*. On the other hand, if there are one or more differences in attributes other than the primary attribute (i.e., *secondary attributes*), *secondary adjustments* are required. For example, if in addition to having no fireplace, one of the two houses also has no dishwasher, but its mate does, this matched pair is not a pure pair for estimating the primary adjustment for a fireplace. Therefore, in the case of PDA, before the primary adjustment can be estimated, a secondary adjustment must be made to correct for the lack of a dishwasher.

[4]The requirement that secondary adjustments must come from within the data set is needed to develop the equivalence relationship. The notion of restricting secondary adjustment information to the data set used in the model is consistent with regulatory requirements for market-derived adjustments and is theoretically superior to incorporating outside sources such as replacement cost, which we know is not always equal to market value, especially when it pertains to individual components.

[5]Lipscomb and Gray (1990) use a criterion for matching based on the importance of variables as measured by their contribution to R^2 in a regression of sales price on the explanatory variables. See Tchira (1979) and Vandell (1979) for other matching criteria.

[6]Technically, pairs are never perfectly matched. No two properties can be identical because of the location issue. We use the term "rarely" to recognize that in practice some differences are so trivial they can be ignored.

References

Cannaday, R., How Should You Estimate and Provide Market Support for Adjustments in Single Family Appraisals?, *Real Estate Appraiser and Analyst*, Winter 1989, 55:4, 43–54.

Colwell, P., R. Cannaday and C. Wu, The Analytical Foundations of Adjustment Grid Methods, *AREUEA Journal,* Spring 1983, 11:1, 11–29.

Gau, G. and D. Kohlhepp, Alternative Estimation Methods for Reduced-Form Price Equations Under Conditions of Multicollinearity, *AREUEA Journal*, Spring 1978, 6:1, 50–69.

———, Alternative Estimation Methods for Reduced-Form Price Equations Under Conditions of Multicollinearity: Reply, *AREUEA Journal*, Fall 1979, 7:3, 437–41.

Gilley, O. W. and R. K. Pace, A Hybrid Cost and Market-Based Estimator for Appraisal, *Journal of Real Estate Research*, Fall 1990, 5:1, 75–88.

Hamilton, L. C., *Regression with Graphics: A Second Course in Statistics,* Pacific Grove, Calif: Brooks/Cole, 1992.

Kang, H. B. and A. K. Reichert, An Empirical Analysis of Hedonic Regression and Grid-Adjustment Techniques in Real Estate Appraisal, *AREUEA Journal*, Spring 1991, 19:1, 70–91.

Lipscomb, J. B. and J. B. Gray, An Empirical Investigation of Four Market-Derived Adjustment Methods, *Journal of Real Estate Research,* Fall 1990, 5:1, 53–66.

Rosen, S., Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *Journal of Political Economy,* January–February 1974, 82:1, 34–55.

Tchira, A., Comparable Sales Selection – A Computer Approach, *Appraisal Journal,* Spring 1979, 47:1, 86–98.

Vandell, K., Alternative Estimation Methods for Reduced-Form Price Equations Under Conditions of Multicollinearity: A Comment, *AREUEA Journal,* Fall 1979, 7:3, 427–36.

———, Optimal Comparable Selection and Weighting in Real Property Valuation, *Journal of the American Real Estate and Urban Economics Association*, Summer 1992, 19:2, 213–39.